Available online at www.sciencedirect.com
ScienceDirect
Procedia Economics and Finance 37 (2016) 214 – 220
Procedia Economics and Finance
www.elsevier.com/locate/procedia

FIFTH INTERNATIONAL CONFERENCE ON MARKETING AND RETAILING (5TH INCOMaR) 2015

Customer Voice Retaliation (CVR) Construct Verification: A Rasch Analysis Approach

Nor Irvoni Mohd Ishar[a]* Rosmimah Mohd Roslin[b]

[ab]*Faculty of Business and Management, Universiti Teknologi MARA,Shah Alam 40450 Malaysia*

Abstract

The study on customer complaining behaviour has received significant attention over the past decades. Generally, customers expect a fair treatment for the investments made in a relationship with the organisation. Hence, perceived unfairness will create an impression that they have been betrayed, and thus motivate them to complain as a way of expressing dissatisfaction. In extreme cases, they might opt for aggressive complaining to compensate unfairness. In this study, the term aggressive complaining will be replaced by customer voice retaliation (CVR). Based on previous literatures, a framework was developed to measure CVR which consist of three constructs. The aim of this study is to verify the construct for a new CVR framework. Rasch analysis was used to examine reliability for both respondents and items. It should give us the list of items that should be included in measuring the CVR constructs. Sample used for the study were 27 for pre-testing, and 66 for the pilot study. Respondents consisted of subscribers who had experienced dissatisfaction, and had to some extent performed complaining behaviour. From the pre-test analysis, item polarity indicates that all items were measuring in the same direction. Similarly, the summary statistic from the pilot study indicated that item reliability and item separation was 0.95 and 4.39 respectively, while for person reliability and person separation was 0.95 and 4.6 respectively. However, the result from the pilot test for model Unidimensionality suggested the existence of a secondary dimension which was a possible contributor to multicollinearity problem. This indicates that items for measuring CVR needs to be reviewed and instrument construct validity call for further refinement.

Peer-review under responsibility of Faculty of Business Management, Universiti Teknologi MARA

Keywords: customer complaining behavior, customer voice retaliation , negative emotional experience , Rasch Analysis, service experience dissatisfaction,

* Corresponding author. Tel.: +603 3258 5074; fax: +603 3258 5000.
E-mail address: irvoni@salam.uitm.edu.my / irvoni.uitm@gmail.com

Peer-review under responsibility of Faculty of Business Management, Universiti Teknologi MARA
doi:10.1016/S2212-5671(16)30116-2

1. Introduction

The Economic Transformation Programme (ETP) initiated by the government envisaged that Malaysia will be a developed nation by 2020. A developed economy as envisioned by the Malaysian government will depend extensively on the service sector. As such, the service sector needs to be top notch, quality oriented and more importantly able to deliver what is promised.

Inability to deliver what was promised is detrimental to any service providers. Customers not only will complain, but will also opt for more aggressive complaining behaviours such as voice retaliation. In the past, amongst the aggressive behaviours that have been studied includes create loss, vandalism, trashing, stealing, negative word of mouth, and personal attack (Huefner & Hunt, 2000). However, specific behavior on aggressive voice or voice retaliation has yet to be explored. In this regard, a model that relates to the handling of complaints that are aggressive in nature is likely to enhance the service sector performance and ultimately improve the standing of the economy. Unfortunately, to the authors' current knowledge, such model has yet to be established. This is particularly so for Malaysia because it has the tendency to be viewed as confrontational (Ndubisi & Ling, 2005). Hence, studies that display such attribution may not be regarded as providing significant inputs for marketing. This view however should change as studies on retaliatory behaviours in business settings could enhance our understanding on how to effectively deal with such behaviours. In order to advance our comprehension on CVR, it is necessary to have an in-depth understanding of why, and how customers respond to dissatisfying service experiences, and whether complaints triggered by such experiences will eventually lead to retaliation by the customers. This is still an area that has not been dealt with comprehensively in the service marketing literature. Therefore, this research attempts to develop a model for handling CVR using Rasch analysis. Rasch will help to scrutinize the constructs, and to verify whether items used in the instruments are fit for measuring CVR.

2. Rasch Measurement Model

Social science researchers have been applying the classical test theory (CTT) and use likert rating scale as a way to collect data for measuring latent trait in their studies. In the past, the use of ordered scale between 1 (very low) to 5 (very high), or rating of between 1 (strongly disagree) to 5 (strongly agree) is a common practice among them. Respondents were asked to rate their responses to the given statements that only indicated the rank of order of the different categories. The label used for ranking the categories does not indicate the extent to which level is greater than the other level (Bond & Fox, 2007). This is because data placed in ranking order are not linear, but is continuum in nature, and it does not have an equal interval which is crucial for statistical analysis (Jamieson, 2004). Therefore, to go beyond this ranking, interval, or ratio scales are needed.

Recently, the application of Rasch Measurement Model has gained considerable attention among the social scientist. It has been applied in various area of studies (Battisti, Nicolini, & Salini, 2005; Brentari & Golia, 2008; Hashim, Nor Irvoni, Rashid, & Masodi, 2012; Nor Irvoni & Mohd Saidfudin, 2012). In contrast to CTT, Rasch analysis is probabilistic and inferential, and focuses on the pattern of item responses that stipulates the interaction between a person and an item based on a mutual latent trait (Nor Irvoni & Mohd Saidfudin, 2012). It predicts the likelihood of how a person of different ability level for a particular trait should respond to an item of a certain level of difficulty. The probability of success depends on the difference between the ability of the person and the difficulty of the item (Bond & Fox, 2007). Rasch has the ability to transform ordinal data into interval.

Rasch Model assumes that the item difficulty is the attribute that is influencing the person responses while the person ability is the attribute that is influencing the item difficulty estimates (Linacre, 2010). The Rasch Model theorem assumption is based on two fundamental expectations:

i. A person who is more capable has a greater likelihood to answer all items correctly; and
ii. An easy item is more likely to be answered correctly by all persons.

One important feature of Rasch is Fit statistics. It allows researchers to visualise whether the data used is feasible.

Specifically for this study, more retaliatory subscribers should be more likely to endorse items of greater difficulty than the least retaliatory subscribers. It includes the infit and outfit (mean square and standardized values) of the items and persons. The term fit (Green & Frantom, 2002) refers to “infit” (weighted by the distance between the person position and item difficulty) and “outfit” (an unweighted measure). Between the two, outfit is more sensitive to extreme responses compared to the infit.

Since Rasch takes into account the relationship between the item difficulty and person’s ability, it is therefore desirable to look at the item fit as well as the person fit. Item fit refers to an index which implies the functionality of the item, and the latter refers to an index which signifies the responses of an individual. If the data fits the Rasch model, then the expected values of the standardized values, and mean square fit indices are 0 and 1 respectively. A misfit item means that the particular item is either too easy, or too difficult for the person to endorse. It could also mean that the item is not really measuring the desired latent trait. Similarly, any erratic or irregular response could be a sign of a misfit. In order to verify the fit and misfit for persons or items, the following goodness of fit criteria must be satisfied (Nor Irvoni & Mohd Saidfudin, 2012) :

i. Point Measure Correlation (PMC), $0.4 < x < 0.8$
ii. Mean Square (Infit & Outfit), $0.5 < y < 1.5$
iii. Z standard (Infit & Outfit), $-2.0 < Z < +2.0$

3. Method

A. The Instrument

The instrument for this study consists of 74 items, adapted from various sources. The instrument is divided into three sections which basically are the main constructs understudy. The constructs are customer voice retaliation (CVR), negative emotional experience (NEMEx), and dissatisfied service experience attribution (DExSA). None of the items are negatively worded as Rasch Analysis is able to detect any irregular pattern by the respondents through the value indicated by PMC.

B. Pre-test

The questionnaire was pre-tested in two stages. The first stage was done with three content experts from the domain area, three Rasch Measurement experts, and three industry practitioners. The aim is to validate the content and face validity of the instrument. The experts helped to identify unclear terms, and poorly worded questions, and the questionnaire was revised accordingly.

The second stage of pre-testing involves distributing the instrument to a small convenience sample of 27 respondents, in three separate sessions. Inclusion criteria included subscribers who have experienced service dissatisfaction, and have performed some form of complaining behaviour.

From the pre-test, reliability analysis using Cronbach’s alpha (KR-20), was performed to ensure that respondents were responding consistently to items. With Cronbach’s alpha value of 0.96, the survey instrument can be considered as having excellent internal consistency (Fisher, 2007). Additional test was also conducted using the PMC. Results showed that none of the items gave any negative values. This indicates that all items are measuring in the right same direction. With these analysis, and feedbacks from the expert panels, the final questionnaire was developed for pilot study.

C. Pilot Study

The pilot test was conducted on 66 respondents who are subscribers of the local Telco service providers. It examined the mediating effect of negative emotional experience on subscribers’ dissatisfied service experience, and voice retaliation among the subscribers in the Klang Valley. A Mall intercept technique was used at selected locations. The purpose of the pilot test is to further confirm the reliability of the items, and to ensure that the measures, and meanings of the constructs are suitable for the study. Hence, instrument construct verification has taken place through a more rigorous process of reliability test, and quality checks on instrument was done via the analysis of Unidimensionality and item dependency, and goodness of fit test.

4. Construct Verification Analysis

4.1 Reliability Indices

In this preliminary analysis, the Rasch item-person statistics in Table 1 are derived from the pilot data to examine the fit of the data to the Rasch model. In particular, Rasch diagnostics check the interdependency among the persons' parameters and items' parameters, compare the estimated probabilities, and the observed proportions for the items, and analyses the persons' and items' residual structures (Battisti, Nicolini, & Salini, 2010). This process is crucial to see the fit of the measures of each item to allow for further analysis.

Table 1: Summary of Fit Statistics for Customer Voice Retaliation

(N=66, i=74)	Item – Person Interaction			
	Item Location		Person Location	
Mean	0.00		-0.28	
SD	0.58		0.52	
MnSq (Infit & Outfit)	1.04	1.03	1.0	1.03
z-std (Infit & Outfit)	0.1	0	-0.3	-0.1
Maximum Measure	1.24	(SE=0.15)	1.0	(SE=0.13)
Minimum Measure	-0.91	(SE=0.10)	-1.24	(SE=0.10)
Reliability Indices				
Separation	4.60	(very good)	4.39	(very good)
Reliability	0.95	(excellent)	0.95	(excellent)
Model error	0.11	(very good)	0.11	(very good)
Cronbach's Alpha (KR-20)	0.96 (excellent)			

A total of 4881 data points from 66 respondents on 74 items was analyzed using WINSTEPS 3.7.23 software. Items difficulty, and person ability or person measure location, are expressed in logits through the transformation of raw score (ordinal type scale) percentage into its probability of success-to-failure ratio or odds which was then converted to its natural log known as the logit unit. Rasch provides indicators of how well the items fit within the underlying construct. The result yielded a Chi-square value of 14083 with 4738 degree of freedom. The test raw score Cronbach's Alpha registers a reliability of 0.96, which allows for further analysis of the instrument.

The goodness of fit of the survey instrument is described by the precision or errors in the item difficulty estimates and person ability estimates, person fit, item fit, and reliabilities of item and person estimates. Table 1 shows the summary of items estimates with the mean defaulted at 0 logit. Item reliability is excellent at +0.95 logits on a 0 – 1 scale similar to interpreting Cronbach's Alpha, and is transformed to item separation index of 4.6, indicating a very good item range (Fisher, 2007). Hence, the order of items is replicable across comparable cohorts.

The subscribers' ability estimates mean of –0.28 logits in Table 1 indicates that the item is comparatively difficult for the respondents to endorse. The maximum item measure is at +1.24logits (SE=0.15) which is slightly higher than the maximum measure for person. Although the item logitmax is slightly higher than person logitmax, it still provides a good targeting between items and persons.

The instrument is capable of yielding a very good person separation of 4.39, indicating there is very good separation of measures along the scale, compared with the errors of measurement, which are comparatively smaller (.11). This implies that the power of the tests of fit of the model was very good. The Person Infit MnSq value is at the ideal 1, and z-std value that is close to 0 (–0.3 logits), gives an indication of the goodness of fit of the instrument, and that it is measuring what is to be measured. In short, the data fits the measurement model.

4.2 Data Quality Control

A. Construct Unidimensionality and Item Dependency

Unidimensionality is a crucial element in determining construct validity. In Rasch, it can be identified using the Principal Component Analysis as depicted in Table 2.

Table 2: Standardized Residual Variance (in eigenvalue units)

Description	Empirical (%)	Model
Variance explained by measures	42.6%	43.0%
Unexplained variance	57.4%	57.0%
Unexplained variance in 1st contrast	8.5%	

To satisfy Unidimensionality, items in the instruments must measure the same composite of abilities - the customer's voice retaliation. As indicated in Table 2, the Rasch PCA of residuals yields the raw variance explained by measure of 42.6% which is very close to the variance expected by the model (43%). It fulfilled the minimum Unidimensionality threshold of 40% is an indication of a strong measurement dimension (Conrad, Conrad, Dennis, Riley, & Funk, 2011). The unexplained variance in the first factor of 8.5% indicates the instrument as good (Fisher, 2007).

Item dependency or multicollinearity can also be identified from the largest standardized residual correlation table. It highlights item pairs that are locally dependent. Locally dependent items pairs which are highly correlated (> 0.7) can be considered as redundant. Hence, needs to be rephrased or deleted. The local dependence test for the largest standardized residuals correlation (in Table 3) suggests that out of 74 items, 8 items (4 pairs) is a possible contributor to multicollinearity problems as it breached the 0.7 limit (Ferketich, 1991). Therefore, it is suggested for those 8 items to be revisited for deletion, or rephrasing. However, further checks on the items found that none of the item violates the goodness of fit criteria (PMC, Outfit MnSq and Outfit z-std). Therefore, all 74 items will be included for further analysis on fit.

Table 3: Largest Standardize Residual Correlation

Correlation	Item	Item
0.78	30 SOCi_6	34 SOCi_10
0.77	59 SW_1	61 SW_3
0.76	57 VIN_3	63 COL_1
0.74	59 SW_1	60 SW_2
0.73	68 PAt_2	70 PAt_4
0.69	39 rag_1	41 rag_3
0.68	57 VIN_3	64 COL_2
0.68	1 ATM_1	26 SOCi_2
0.68	60 SW_2	61 SW_3
0.67	15 PRO_1	16 PRO_2

B. Goodness of Fit Test

Rasch goodness of fit test provides an indication of how well an item fit the model. It helps researcher to identify misfit items for further refinement of a test instrument. An item is a misfit when it violates the goodness of fit criteria for PMC, Outfit Mn Sq, and Outfit z-std values. Generally, in Table 4, all items register a positive PMC, and a small measurement error. In this case, one item (PAt_1) is considered as misfits, hence a candidate for deletion. Scrutinization conducted on the item revealed that its elimination will improve item reliability. Further, its elimination will not disrupt the content validity of construct understudy. Hence, item 67 (PAt_1) will be eliminated from the instrument. As a result, only 73 items will be used in the final survey.

Table 4: Item Statistics – Misfit Order

Item No	Measure	Model SE	INFIT		OUTFIT		PMC	Item
			MnSq	zstd	MnSq	zstd		
63	0.69	0.13	1.94	4.2	1.79	3.7	0.43	COL_1
67	0.51	0.12	1.71	3.4	1.85	4.0	0.33	PAt_1
64	0.63	0.12	1.83	3.8	1.76	3.6	0.45	COL_2
70	0.74	0.13	1.75	3.5	1.64	3.1	0.51	PAt_4
57	0.89	0.13	1.62	2.9	1.47	2.3	0.47	VIN_3
68	0.84	0.13	1.51	2.5	1.47	2.3	0.53	PAt_2
6	-0.51	0.10	1.41	2.4	1.50	2.8	0.25	ATM_6
4	-0.34	0.10	1.32	1.9	1.39	2.2	0.22	ATM_4
59	-0.50	0.10	1.36	2.1	1.37	2.2	0.54	SW_1
1	0.21	0.11	1.21	1.3	1.25	1.4	0.19	ATM_1
22	0.84	0.13	1.24	1.3	1.19	1.0	0.39	CORS_2
3	-0.29	0.11	1.14	0.9	1.18	1.1	0.32	ATM_3
33	0.8	0.13	1.09	0.6	1.09	0.6	0.35	SOCi_9
				BETTER FITTING OMITTED				

5. Conclusion and Recommendation

An attempt at refining a research instrument using the Rasch method of analysis yielded more robust constructs for measuring Customer Voice Retaliation (CVR). It promotes the development of more powerful substantial theories of constructs as it additionally asks for the characterization of different levels of the latent variable (Salzberger, 2000). Indeed, items that are postulated to measure CVR are exposed to more rigorous exercise in ensuring that the instrument yielded measures that have better fit and therefore are more likely to result in reliable and valid findings.

Table 5: Cleaned CVR-73 Instrument Construct Properties

Description	Item	Person
Reliability	0.96 (0.95)	0.95 (0.95)
Separation	4.63 (4.60)	4.40 (4.39)
Infit MnSq SD	0.32 (0.32)	0.42 (0.43)
Mean Error	0.11 (0.11)	0.11 (0.11)
PCA Variance explained by measures	42.9% (42.6%)	
Unexplained variance in 1st contrast	8.5% (8.5%)	

A better fit instrument is constructed when all parameters are met. Improvements on instrument are reanalyzed for reliability, Unidimensionality and fit. Table 5 are the cleaned values of 73 items together with the original CVR items in bracket indicating a slight improvements in item reliability reading (+0.96 logits) and other fit indices remain unchanged. This evidence provides support that Rasch analysis offers a powerful tool for putting all these conjectures to the test in the context of measurement (Salzberger, 2000). Subsequently, the instrument can be adapted for future research in other similar settings. Other than that, further checks on the category functioning of the six-point Likert rating scale used in the study should be re-examined.

Acknowledgement

The authors are grateful to all involved at the Faculty of Business & Management, Universiti Teknologi MARA, and to the Ministry of Higher Education, Malaysia for the support given in carrying out this study.

References

Battisti, F. De, Nicolini, G., & Salini, S. (2005). *The Rasch Model to Measure Service Quality*.

Battisti, F. De, Nicolini, G., & Salini, S. (2010). The Rasch Model in Customer satisfaction Survey Data. *Quality Technology & Quantitative Management*, *7*(1), 15–33.

Bond, T. G., & Fox, C. M. (2007). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates.

Brentari, E., & Golia, S. (2008). Measuring job satisfaction in the social services sector with the Rasch model. *Journal of Applied Measurement*, *9*(1), 45–56.

Conrad, K. J., Conrad, K. M., Dennis, M. L., Riley, B. B., & Funk, R. (2011). *Validation of the Substance Problem Scale to the Rasch Measurement Model, GAIN Methods Report 1.2*.

Ferketich, S. (1991). Focus on psychometrics. Aspects of item analysis. *Research in Nursing & Health*, *14*(2), 165–168.

Fisher, W. J. (2007). Rating Scale Instrument Quality Criteria . *Rasch Measurement Transaction*, *21*(1), 2007.

Green, K., & Frantom, C. (2002). Survey development and validation with the Rasch model. In *International Conference on Questionnaire Development, Evaluation, and Testing* (pp. 1–30).

Hashim, N., Nor Irvoni, M. I., Rashid, W. E. W., & Masodi, M. S. (2012). Personality Traits, Work-Family Conflict and Job Satisfaction: Items Validity using Rasch Measurement Approach. *Procedia - Social and Behavioral Sciences*, *65*(ICIBSoS), 1013–1019.

Huefner, J. C., & Hunt, H. K. (2000). Consumer Retaliation As A Response to Dissatisfaction. *Journal of Consumer Satisfaction Dissatisfaction and Complaining Behavior*, *13*(1), 61 – 82.

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, *38*(12), 1217–8.

Linacre, J. M. (2010). Predicting responses from Rasch measures. *Journal Of Applied Measurement*, *11*(1), 1–10.

Ndubisi, N. O., & Ling, T. Y. (2005). Complaint behaviour of Malaysian consumers. *Management Research News*, *29*(1/2), 65–76.

Nor Irvoni, M. I., & Mohd Saidfudin, M. (2012). Students' Perception towards Quality Library Service Using Rasch Measurement Model. In *2012 International Conference of Innovation, Management and Technology Research (ICIMTR2012)* (pp. 668 – 672).

Salzberger, T. (2000). An Alternative Way of Establishing Measurement in Marketing Research – Its Implications for Scale Development and Validity, 1111–1117.